Estimation of a system of national accounts: implementation with mathematica

Temel, Tugrul
Development Research Institute (IVO), Tilburg University

08. December 2011

Estimation of a System of National Accounts: Implementation with *Mathematica*

Tugrul Temel
Development Research Institute (IVO)
Tilburg University, Tilburg
The Netherlands
t.temel@uvt.nl

December 15, 2011

Abstract

This study implements *Mathematica* to estimate a system of national accounts. The estimation methods applied are portrayed in Danilov and Magnus (2008), including the Bayesian estimation, restricted and unrestricted least-squares estimation and best linear unbiased estimation. Operationalizing these methods in the *Mathematica* environment is the main contribution of the current study. In light of the United Nations' efforts aimed to standardize across countries the compilation of national accounts, the *Mathematica* codes developed here should provide an important tool both for the estimation of unrealized or unavailable national accounts data and for conducting cross-country and within-country macroeconomic policy analysis.

Key words: System of national accounts; Social Accounting Matrix; Bayesian estimation; Least-squares estimation; Best linear unbiased estimation; Linear programming

1 Introduction

This study implements *Mathematica* to estimate a system of national accounts (SNA). The estimation methods applied are portrayed in Danilov and Magnus (2008), including the Bayesian estimation, restricted and unrestricted least-squares estimation and best linear unbiased estimation. Operationalizing these methods in the *Mathematica* environment is the main contribution of the current study. In light of the United Nations' efforts aimed to standardize across countries the compilation of national accounts, the *Mathematica* codes developed here should provide an important tool both for the estimation of unrealized or unavailable national accounts data and for conducting cross-country and within-country macroeconomic policy analysis. The *Mathematica* codes should benefit the most statistics organizations responsible for the compilation and updating of national accounts and policy-making bodies drawing on national accounts data.

The study is organized into five sections. Following the Introduction, Section 2 formulates a data estimation problem drawing on an example system of national accounts. Section 3 describes four estimation methods. Section 4 describes the implementation of the computational algorithm developed and presents the estimations concerning the example SNA. Finally, Section 5 concludes the paper with some remarks on the efficiency of *Mathematica* for solving large linear systems.

2 An example system of national accounts[1]

2.1 Set-up

Consider an example SNA illustrated in Figure 1. The SNA consists of two sub-accounts: supply and use accounts (SUA) and integrated economic accounts (IEA), and is fully characterized by 10 variables. Given a benchmark (or reference) SNA at period $t = 0$ and 4 variables known with precision at period $t = 1$ (see Figure 2), the goal is to estimate the remaining 6 variables for $t = 1$. To do that, we utilize two additional pieces of information. First, five indicator ratios are constructed drawing on expert knowledge about the long-run behavior of the economy concerned. Second, six linear restrictions (or identities) are introduced using macro-accounting relations among the variables in the SNA (see Figure 3).

Below, we describe the set-up using mathematical notations.

- Notations

  $\tilde{x}_i^t$ denotes time t value of the i^{th} variable with $i = 1, 2, ..., \tilde{n}$ and $t = 0, 1$.

  $\tilde{x}^t = (\tilde{x}_1^t, \tilde{x}_2^t, ..., \tilde{x}_{\tilde{n}}^t)'$ is a column vector of $\tilde{n}$ variables at time t.

  $f_k(\tilde{x}^t) = 0$ defines the k^{th} identity as a function of $\tilde{x}^t$.

- Assumptions

  1) $\tilde{x}_i^t \sim N(\mu_i^t, \sigma_i^2)$ for all i.

  2) $R_{ij}^0 = R_{ij}^1$ for $\tilde{m}_1$ prior indicator ratios where $R_{ij}^0 = \left(\frac{\tilde{x}_i^0}{\tilde{x}_j^0}\right)$ is a prior indicator ratio evaluated using benchmark data. Reliability levels and reliability coefficients for $\tilde{p}$ data observations made at $t = 1$ and for $\tilde{m}_1$ indicator ratios are set using prior information about the quality of available data: {Fixed = 0, Strong = 0.01, High = 0.03, Medium = 0.06, Low = 0.12, Poor = 0.24}.

[1] I would like to thank Jan van Tongeren for letting me use this example system of national accounts.

- Available data, priors and restrictions

  1) At $t = 0$, benchmark data are available on $\tilde{n}$ variables: $\tilde{x}^0 = (\tilde{x}^0_1, \tilde{x}^0_2, ..., \tilde{x}^0_{\tilde{n}})'$.

  2) At $t = 1$, data are available on $\tilde{p}$ variables: $(\tilde{x}^1_1, \tilde{x}^1_2, ..., \tilde{x}^1_{\tilde{p}})'$ with $\tilde{p} < \tilde{n}$.

  3) At $t = 1$, $\tilde{m}_1$ indicator ratios are available

  4) At $t = 0, 1$, linear restrictions $f_k(\tilde{x}^t) = 0$ hold for all $k = 1, 2, ..., \tilde{m}_2$

Technically speaking, the objective is to estimate the posterior mean and variance of variables in vector $\tilde{x}^1$, on some of which data are available with precisions ($\tilde{p}$), on some prior indicator ratios are available with precisions ($\tilde{m}_1$), and on some prior information is available with precisions ($\tilde{m}_2$).

2.2 Data and prior information

In what follows, we translate the data and information given in Figures 1-3 into mathematical format so that one can link the example SNA to the estimation methods formally presented in the following sections. Suppose that at time $t = 0$ a benchmark (or reference) data set is available for a vector of $\tilde{n} = 10$ latent variables, denoted by:

$$\begin{aligned} \tilde{x}^0 &\equiv (\tilde{x}^0_1, \tilde{x}^0_2, \tilde{x}^0_3, \tilde{x}^0_4, \tilde{x}^0_5, \tilde{x}^0_6, \tilde{x}^0_7, \tilde{x}^0_8, \tilde{x}^0_9, \tilde{x}^0_{10})' \\ &= (P^0, M^0, I^0, K^0, X^0, C^0, Y^0, R^0, S^0, B^0)' \\ &= (100, 90, 40, 30, 70, 50, 60, 60, 10, 20)'. \end{aligned}$$

At time $t = 1$, data are available only for 4 variables ($\tilde{p} = 4$):

$$\begin{aligned} (\tilde{x}^1_1, \tilde{x}^1_2, \tilde{x}^1_4, \tilde{x}^1_5)' &\equiv (P^1, M^1, K^1, X^1)' \\ &= (107.12, 93.64, 32.14, 74.98)' \end{aligned}$$

Furthermore, the benckmark values of the following 5 prior indicator ratios ($\tilde{m}_1 = 5$) are assumed to remain the same over the period from $t = 0$ to $t = 1$, implying that the economy has been in the state of equilibrium during that period.

$$\begin{aligned} (\frac{I^0}{P^0}) &= \frac{40}{100} = 0.4 \\ (\frac{K^0}{P^0 + M^0}) &= \frac{30}{190} = 0.158 \\ (\frac{M^0}{P^0}) &= \frac{90}{100} = 0.9 \\ (\frac{C^0}{R^0}) &= \frac{50}{60} = 0.833 \\ (\frac{X^0}{P^0}) &= \frac{70}{100} = 0.7 \end{aligned}$$

As a last piece of information, we assume that the following 6 linear restrictions ($\tilde{m}_2 = 6$) hold across all t, reflecting the basic macro-accounting relations among the $\tilde{n}$ variables:

$$\begin{aligned}
Y^t - P^t + I^t &\equiv 0 \\
S^t - R^t + C^t &\equiv 0 \\
B^t - M^t + X^t &\equiv 0 \\
Y^t - R^t &\equiv 0 \\
K^t - S^t - B^t &\equiv 0 \\
P^t + M^t - I^t - C^t - K^t - X^t &\equiv 0.
\end{aligned}$$

2.3 The system of linear equations

Data available ($\tilde{p} = 4$) at $t = 1$ are expressed as:

$$\begin{aligned}
\tilde{D}_1 \tilde{x}^1 &= \tilde{d}_1 \text{ where} \\
\tilde{D}_1 &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
\tilde{x}^1 &= (P^1, M^1, I^1, K^1, X^1, C^1, Y^1, R^1, S^1, B^1)' \\
\tilde{d}_1 &= (107.12, 93.64, 32.14, 74.98)'
\end{aligned}$$

The "linearized" indicator ratios ($\tilde{m}_1 = 5$) assumed to hold for all t are expressed as

$$\begin{aligned}
\tilde{A}_1 \tilde{x}^t &\equiv \tilde{h}_1 \text{ where} \\
\tilde{A}_1 &= \begin{bmatrix} -0.4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.158 & -0.158 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.833 & 0 & 0 \\ -0.7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
\tilde{x}^t &= (P^t, M^t, I^t, K^t, X^t, C^t, Y^t, R^t, S^t, B^t)' \\
\tilde{h}_1 &= (0, 0, 0, 0, 0)'
\end{aligned}$$

The linear restrictions ($\tilde{m}_2 = 6$) are:

$$\begin{aligned}
\tilde{A}_2 \tilde{x}^t &\equiv \tilde{h}_2 \text{ for all } t, \text{ where} \\
\tilde{A}_2 &= \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \\
\tilde{x}^t &= (P^t, M^t, I^t, K^t, X^t, C^t, Y^t, R^t, S^t, B^t)' \\
\tilde{h}_2 &= (0, 0, 0, 0, 0, 0)'
\end{aligned}$$

2.4 The modified system of linear equations

Owing to the numerator $(P^t + M^t)$ of the 2^{nd} indicator ratio above, we define a composite variable $Z^t \equiv (P^t + M^t)$ where $Z^0 = 190$ and $Z^1 = 200.76$. The introduction of this composite variable requires some modifications in the linear system described in section (2.2). The first modification takes place in $\tilde{D}_1\tilde{x}^1 = \tilde{d}_1$ as follows:

$$
\begin{aligned}
D_1 x^1 &= d_1 \text{ where} \\
D_1 &= \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \\
x^1 &= (P^1, M^1, I^1, K^1, X^1, C^1, Y^1, R^1, S^1, B^1, Z^1)' \\
d_1 &= (107.12, 93.64, 32.14, 74.98, 200.76)'
\end{aligned}
$$

The second modification takes place in $\tilde{A}_1\tilde{x}^t \equiv \tilde{h}_1$ as follows:

$$
\begin{aligned}
A_1 x^t &\equiv h_1 \text{ where} \\
A_1 &= \begin{bmatrix}
-0.4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.158 \\
-0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.833 & 0 & 0 & 0 \\
-0.7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \\
x^t &= (P^t, M^t, I^t, K^t, X^t, C^t, Y^t, R^t, S^t, B^t, Z^t)' \\
h_1 &= (0, 0, 0, 0, 0)'
\end{aligned}
$$

The third modification takes place in $\tilde{A}_2\tilde{x}^t \equiv \tilde{h}_2$ by introducing the new identity $Z^t \equiv (P^t + M^t)$:

$$
\begin{aligned}
A_2 x^t &\equiv h_2 \text{ where} \\
A_2 &= \begin{bmatrix}
-1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\
1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \\
x^t &\equiv (P^t, M^t, I^t, K^t, X^t, C^t, Y^t, R^t, S^t, B^t, Z^t)' \\
h_2 &= (0, 0, 0, 0, 0, 0, 0)'
\end{aligned}
$$

Our goal is to solve the following system of equations using the estimation methods introduced in Section 3.

$$\begin{aligned}
D_1x^1 &= d_1 \text{ with } \; d_1 \mid x^t \sim N_p(D_1x^t, \Sigma_1) \\
A_1x^1 &\equiv h_1 \text{ with } \; A_1 \sim N_{m_1}(h_1, H_1) \\
A_2x^1 &= h_2 \text{ with } A_2 \sim N_{m_2}(h_2, H_2) \text{ (asy.)} \\
\text{where } \; p &= (\tilde{p}+1) = 5 \\
n &= (\tilde{n}+1) = 11 \\
m_1 &= \tilde{m}_1 = 5 \\
m_2 &= (\tilde{m}_2+1) = 7 \\
m &= (m_1+m_2) = 12
\end{aligned}$$

3 Estimation Methods

The reader is referred to Danilov and Magnus (2008) for a detailed description of the estimation problems stated below. Although they are equivalent and all yield the same estimations, the performance of their computerized solution algorithms differ substantially depending on the size and sparsitiy of the linear system concerned.

3.1 Bayesian estimation

Assume (i) $d_1|x^1 \sim N_p(D_1x^1, \Sigma_1)$ where $D_{1,(p,n)}$ has full row-rank and Σ_1 is positive definite (hence non-singular); (ii) $Ax^1 \sim N_m(h, H)$ where $A = (A_1 : A_2)$, a column vector $h = (h_1, h_2)$, a block diagonal matrix $H = (H_1, H_2)$ with H_1 associated with A_1 and H_2 with A_2; (iii) A has full row-rank and H may be singular. If $m < n$, let L be a semi-orthogonal $(n, n-m)$ matrix such that $L^TL = I_{n-m}$ and $AL = 0$, and assume that the identifiability condition $r(A) + r(D_1L) = n$ is satisfied. Then the posterior distribution of x^1 is given by $x^1|d_1 \sim N_n(\mu, V)$ with

$$\begin{aligned}
V &= A^+HA^{+'} - A^+HA^{+'}D_1^{'}\Sigma_0^{-1}D_1A^+HA^{+'} + CKC' \\
\mu &= A^+h - (A^+HA^{+'} + CK)D_1^{'}\Sigma_0^{-1}(D_1A^+h - d_1) \\
\text{where } \; A^+ &= A^{'}(AA^{'})^{-1} \quad \text{"the Moore-Penrose inverse"} \\
\Sigma_0 &= \Sigma_1 + D_1A^+HA^{+'}D_1^{'} \\
C &= I_n - A^+HA^{+'}D_1^{'}\Sigma_0^{-1}D_1 \\
K &= \left\{ \begin{array}{lll} L(L^{'}D_1^{'}\Sigma_0^{-1}D_1L)^{-1}L^{'} & \text{if } \; m < n & \text{(Lemma A2)} \\ 0 & \text{if } m = n & \text{(Lemma A1)} \end{array} \right\}
\end{aligned}$$

(see Theorem 1 in *Magnus, Tongeren and Vos (2000)).* A^+ denotes the Moore-Penrose (MP) inverse of A. All the variables with superscript (+) stand for the MP inverse.

3.2 Restricted least-squares estimation

For estimations in large systems, the least-squares method works better compared to the Bayesian estimation method (Theorem 1). The Bayesian problem above can be equivalently formulated as a restricted least-squares problem:

$$\underset{x}{\text{Minimize}}(d - Dx)'\Sigma^{-1}(d - Dx) \quad \text{subject to} \quad A_2 x = h_2$$

$$\begin{aligned} \text{where } d &= Dx + \epsilon \\ d &\mid x \sim N_{p+m_1}(Dx, \Sigma) \\ \epsilon &\sim N_{p+m_1}(0, \Sigma) \text{ and} \\ d &= \begin{pmatrix} d_1 \\ h_1 \end{pmatrix}; \quad D = \begin{pmatrix} D_1 \\ A_1 \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & H_1 \end{pmatrix}. \end{aligned}$$

From Theorem 36 in Magnus and Neudecker (1999) (p.233), the general solution, x, to this minimization problem is:

$$\begin{aligned} x &= x_0 + N^+ A_2^{'}(A_2 N^+ A_2^{'})^+(h_2 - A_2 x_0) + (I_n - NN^+)q \\ \text{where } x_0 &= N^+ D^{'} \Sigma^{-1} d \\ N &= D^{'} \Sigma^{-1} D + A_2^{'} A_2 \\ q &= \text{an arbirary vector} \end{aligned}$$

Compute x by using:

$$d = \begin{bmatrix} 107.12 \\ 93.64 \\ 32.14 \\ 74.98 \\ 190 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -0.4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.16 \\ -0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.8 & 0 & 0 & 0 \\ -0.7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 41.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 29.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 35.8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 17.6 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

3.3 Unrestricted least-squares estimation

Alternatively, the solution of the following unrestricted minimization problem is also identical to that of the Bayesian estimation of x :

$$\text{Minimize} \begin{pmatrix} d - Dx \\ h_2 - A_2x \end{pmatrix}' \begin{pmatrix} \Sigma + DD' & DA_2' \\ A_2D' & A_2A_2' \end{pmatrix} \begin{pmatrix} d - Dx \\ h_2 - A_2x \end{pmatrix}$$

$$\text{where } d = \begin{pmatrix} d_1 \\ h_1 \end{pmatrix}, \quad D = \begin{pmatrix} D_1 \\ A_1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & H_1 \end{pmatrix}$$

3.4 Best linear unbiased estimation

Best linear unbiased estimation (BLUE) is an alternative method that leads to the same results as restricted least-squares method. Consider the regression model:

$$\text{Minimize } (d - Dx)'\Sigma^{-1}(d - Dx) \quad \text{subject to: } A_2x = h_2$$

where x is a vector of parameters to be estimated. The BLUE estimator of x is given by:

$$\begin{aligned} x &= G^{-1}D'\Sigma^{-1}d + G^{-1}A_2'(A_2G^{-1}A_2')^{-1}(h_2 - A_2G^{-1}D'\Sigma^{-1}d) \\ \text{with variance } V &= G^{-1} - G^{-1}A_2'(A_2G^{-1}A_2')^{-1}A_2G^{-1} \\ \text{where } G &= D'\Sigma^{-1}D + A_2'A_2 \\ d &= \begin{pmatrix} d_1 \\ h_1 \end{pmatrix}, \quad D = \begin{pmatrix} D_1 \\ A_1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & H_1 \end{pmatrix} \end{aligned}$$

4 Implementation

4.1 Algorithm

Mathematica codes for each one of the estimation problems above have been developed using *Mathematica* 8.0. The implementation algorithm applies the following steps.

- STEP 1: Testing the rank condition: $r(A) + r(D_1L) = n$.

  This condition is necessary and sufficient for the existence of a solution. In our example, the rank of A is 11, which is also equal to the number of variables in the system; that is, $r(A) = n = 11$. This shows that the identifiability condition is satisfied but there is one redundant equation in A because $(m - n) = (12 - 11) = 1$. The task is to find out that redundant equation and eliminate it from the system.

- STEP 2: Creating a system of linearly independent equations

  Having applied the Gram-Schmidt method to the set of 7 identities, we find out that 6 identities (excluding $Z^t \equiv P^t + M^t$) are linearly dependent. Hence, dropping any one of the 6 identities from A leads to a system of 11 equations, which then implies that the system at hand is fully identified with rank 11. Here is a sketch of how to perform this task.

  (1) Determine the rank of A with dimension of $(m, n) = (12, 11)$ where $m =$ the number of rows, $n =$ the number of columns.

  (2) $rk(A) = 11$ implies that one of the equations is redundant, which needs to be eliminated from A for the unique solution to exist.

  (3) Apply Gram-Schmidt process to identify the linearly dependent equation. The process would generate a zero row for the dependent equation. Since dependency is a property of a group of equations, not a property of a single equation, Gram-Schmidt process results in a different dependent equation every time we change the order of rows in A. Thus, we obtain 6 alternative systems, each of which has 11 equations and has a non-zero determinant. (Note that a non-zero determinant implies that the system of equations concerned comprises a linearly independent set.)

  (4) There are 6 non-zero determinants. This implies that the final Bayesian estimation should be performed for each one of 6 systems separately and the one that minimizes posterior standard deviation should be used in the final analysis.

- STEP 3: Constructing a variance-covariance matrix

  Due to the elimination of the dependent equation(s) from A, necessary adjustments are made in the vector h and the variance-covariance matrix H.

- STEP 4: Introducing reliability levels and reliability parameters to create an adjusted variance-covariance matrix

  The Bayesian data estimation approach allows for the deviation from the "true" values of the benchmark values of the indicator ratios and the data observations. Reliability levels assigned to each ratio and each one of the 4 observations imply that the "true" values are most likely to lie within the confidence intervals implied by the reliability levels. The concept of reliability bridges the gap between the "true" and "observed" values of a variable. We set reliability levels as: {Fixed (F), Strong (S), High (H), Medium (M), Low (L), Poor (P)}, with arbitrary reliability coefficients of {0, 0.01, 0.03, 0.06, 0.12, 0.24}, respectively. Coefficient of variation, defined by the ratio of standard error to mean, represents reliability coefficient. Given the reliability coefficient and the mean value at $t = 0$ of the variable of interest (see Table 1), we calculate prior standard error of that variable. In the rest of the paper, matrix notation is used for convenience. Estimate the variance-covariance matrices (Σ_1, H) for time $t = 1$ observations and for the indicator ratios, respectively. In the estimation of Σ_1 associated with d_1, the following reliability levels and coefficients are assumed:

Table 1

$Variable$	$R-coeff$	$Mean$	$Prior\ s.e.$	$Prior\ var.$
P^1	M=0.06	107.12	6.43	41.3
M^1	H=0.03	93.64	2.81	7.9
K^1	L=0.12	32.14	3.86	14.9
X^1	H=0.03	74.98	2.25	5.1
Z^1	MH=0.04	200.76	8.03	64.5

Consider, for example, P^1, which is assumed to be observed at the Medium level, with a corresponding reliability coefficient of 0.06. Applying the definition of coefficient of variation $= \frac{se_{P1}}{Mean\ of\ P^1}$ would then yield prior standard error $se_{P^1} = 0.06 * 107.12 = 6.4272$. Thus, the prior variance is 41.3. This operation yields

$$\Sigma_1 = \begin{bmatrix} 41.31 & 0 & 0 & 0 & 0 \\ 0 & 7.89 & 0 & 0 & 0 \\ 0 & 0 & 14.87 & 0 & 0 \\ 0 & 0 & 0 & 5.06 & 0 \\ 0 & 0 & 0 & 0 & 64.49 \end{bmatrix}.$$

The estimation of H associated with A is a bit complex. H_1 corresponding to the indicator ratios is a (5,5) diagonal matrix, elements of which are $\sigma_{ij}^2 B_{ij}^2$, whereas H_2 corresponding to the identities is a (6,6) zero matrix. Table 2 shows how to derive H_1 :

Table 2

$Indicator$ $Ratios$	$R-$ $coeff$	$E(\frac{x_i^t}{x_j^t})$ $= r_{ij}$	$Prior$ $s.e.$	$Prior$ $var\ (\sigma_{ij}^2)$	B_{ij}^2	$\sigma_{ij}^2 B_{ij}^2$
$\frac{x_3^t}{x_1^t} = \frac{I^t}{P^t}$	H=0.03	0.4	0.012	0.00014	10000	1.4
$\frac{x_4^t}{x_{11}^t} = \frac{K^t}{Z^t}$	H=0.03	0.16	0.005	0.00002	36076	0.8
$\frac{x_2^t}{x_1^t} = \frac{M^t}{P^t}$	M=0.06	0.9	0.05	0.0029	10000	29
$\frac{x_6^t}{x_8^t} = \frac{C^t}{R^t}$	L=0.12	0.83	0.1	0.0099	3612	36
$\frac{x_5^t}{x_1^t} = \frac{X^t}{P^t}$	M=0.06	0.7	0.04	0.0018	10000	18

The ratios are assumed to be distributed as $(\frac{x_i^t}{x_j^t}) \sim N_{m_2}(r_{ij}, \sigma_{ij}^2)$, while the linearized ratios as $(x_i^t - r_{ij}x_j^t) \sim N_{m_2}(0, \sigma_{ij}^2 B_{ij}^2)$. The benchmark data x^0 are used to calculate:

$$B_{ij} = \left\{ \begin{array}{ll} \frac{r_{ij}^2}{(1+r_{ij}^2)r_{ij}} x_i + \frac{1}{(1+r_{ij}^2)} x_j & \text{if both } x_i \text{ and } x_j \text{ are available} \\ \frac{x_i}{rij} & \text{if only } x_i \text{ is available} \\ x_j & \text{if only } x_j \text{ is available} \end{array} \right\}$$

- Since $m = n$, Lemma A1 of Theorem 1 applies. The full system is characterized by $(D_1, d_1, \Sigma_1, A, h, H)$, where

$$D_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$d_1 = \begin{bmatrix} 107.12 \\ 93.64 \\ 32.14 \\ 74.98 \\ 190 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 41.3 & 0 & 0 & 0 & 0 \\ 0 & 7.9 & 0 & 0 & 0 \\ 0 & 0 & 14.9 & 0 & 0 \\ 0 & 0 & 0 & 5.1 & 0 \\ 0 & 0 & 0 & 0 & 64.5 \end{bmatrix}$$

$$A = \begin{bmatrix} -0.4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.16 \\ -0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.83 & 0 & 0 & 0 \\ -0.7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 29.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 35.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

4.2 Estimation of the example SNA

The estimation results and the Mathematica codes generating these results are given in Table 3.

5 Concluding remarks

Four estimation methods have been operationalized using *Mathematica 8.0.* An example system of national accounts has been used for illustrative purposes. We have developed a generic *Mathematica* code (translated to C++) for each one of the 4 estimation problems. The codes are applied to very large systems. The power of *Mathematica (C++)* remains to be compared with the SNAER program of Danilov and Magnus (2008).

With the *Mathematica* codes we have developed, national statistics offices responsible for producing quarterly or annual estimations of the SNA would be able to make the estimations with much higher precision relative to the currently practised conventional national accounts estimations. Timely delivery of high precision estimations and the availability of such estimations across countries would pave the way for the policy analysis of cross-country interactions.

References

[1] Abadir, K. M., and Magnus, J. R. (2005). Matrix Algebra. Econometric Exercises 1. Cambridge, UK: Cambridge University Press.

[2] Danilov, D., and Magnus, J.R. (2007). Some equivalences in linear estimation. On-line: http://cdata4.uvt.nl/websitefiles/magnus/paper77b.pdf.

[3] Danilov, D., and Magnus, J. R. (2008). On the estimation of a large sparse Bayesian system: The Snaer program. Computational Statistics & Data Analysis, 52(9), 4203-4224.

[4] Danilov, D., and Magnus, J. R. (2005). Least-squares estimation in large sparse systems. Work in progress.

[5] Golub, Loan. (1983). Matrix Computations.

[6] Huang, C. J., and Crooke, P.S. (1997). Mathematics and Mathematica for economists. Oxford, UK: Blackwell Publishers.

[7] Jing Xiao , Lan Liu , Lirong Xia and Tao Jiang. (2007). Fast Elimination of Redundant Linear Equations and Reconstruction of Recombination-Free Mendelian Inheritance on a Pedigree – Venue: Proc. of 18th Annual ACM-SIAM Symoposium on Discrete Algorithms.

[8] LaMacchia, Odlyzko. 1990. Solving large sparse linear systems over finite fields.

[9] Magnus, J. R., and Neudecker, H. (1999). Matrix di¤erential calculus with applications in statistics and econometrics. New York, USA: John Wiley & Sons Ltd.

[10] Magnus, J. R., van Tongeren, J., and Vos. (2000). National accounts estimation using indicator ratios. Review of Income and Welath, Series 46, No. 3, p-329-350.

[11] Van Tongeren, Jan W., and J.R. Magnus. (2011). Bayesian integration of large SNA data frameworks with an application to Guatemala. Tilburg University, CentER Discussion Paper No. 2011-022

Figure 1: An example system of national accounts

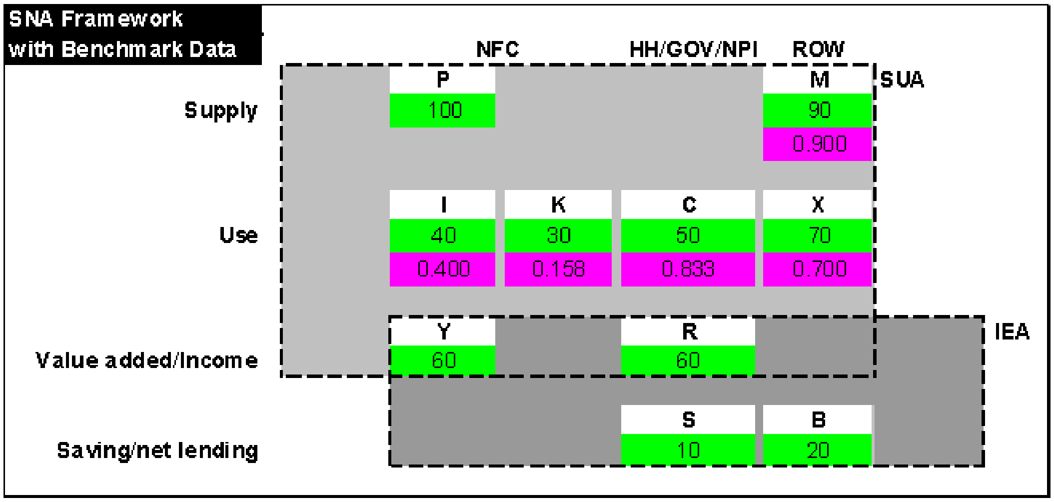

Figure 2: Benchmark and current period data

Variables	Benchmark period t=0	Current period t=1
Output P	100	107.12
Imports M	90	93.64
Intermediate consumption I	40	
Capital formation K	30	32.14
Exports X	70	74.98
Final consumption C	50	
Value added Y	60	
Disposable income R	60	
Savings/Net lending S	10	
External Savings/Net lending B	20	

Figure 3: Indicator ratios and identities (or linear restrictions)

Indicator Ratios	**Definition**	**t=0**
Input-output ratio	I / P	0.400
Investment-supply ratio	K /(P+M)	0.158
Import-output ratio	M / P	0.900
Propensity to consume	C/R	0.833
Exports-output ratio	X / P	0.700

Identities	**Definition**
Value added	Y = P - I
Saving	S = R - C
External deficit to be financed	B = M - X
Income distribution	Y = R
Finance of capital formation	K = S + B
Supply equals use	P + M = I + C + K + X

(*

Table 3: Estimation Outputs

*)

Bayesian Estimation

Variable	Post-mean	Post-se
P	106.181	3.332
M	94.310	2.281
I	42.462	1.791
K	32.047	1.100
X	74.662	1.953
C	51.320	3.168
Y	63.719	2.333
R	63.719	2.333
S	12.400	2.748
B	19.648	2.807
Z	200.491	4.284

Least – squares Estimation

Variable	Est. w/scaling	Est. w/o scaling
P	106.239	107.120
M	94.307	93.640
I	42.494	42.848
K	31.681	32.140
X	74.696	74.980
C	51.681	67.239
Y	63.745	64.272
R	63.658	80.719
S	12.008	13.480
B	19.642	18.660
Z	200.548	200.760

BLUE Estimation

Variable	Post-mean	Post-se
P	106.181	3.331
M	94.309	2.280
I	42.461	1.791
K	32.047	1.100
X	74.662	1.953
C	51.319	3.168
Y	63.719	2.333
R	63.719	2.333
S	12.399	2.748
B	19.647	2.807
Z	200.491	4.282